DOCUMENT RESUME

ED 340 806 UD 028 478

AUTHOR Burke, Art
TITLE Portfolios in Chapter 1 Evaluation and Program Improvement.
PUB DATE Oct 91
NOTE 19p.; Paper presented at the Annual Meeting of the Oregon Educational Research Association (Portland, OR, October 10-11, 1991).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Compensatory Education; Educational Assessment; Elementary Secondary Education; Evaluation Methods; *Federal Programs; *Portfolios (Background Materials); Program Evaluation; *Program Improvement; Student Evaluation
IDENTIFIERS *Education Consolidation Improvement Act Chapter 1; Hawkins Stafford Act 1988

ABSTRACT

This paper addresses methodological and substantive issues regarding the use of portfolios of students' work in Chapter 1 evaluation and program improvement. In evaluations using portfolios desirable characteristics of the evaluation procedure are objectivity, reliability, and validity. The Chapter 1 Evaluation/Program Improvement System (CHEPIS) uses a local evaluation and an annual evaluation. Portfolios are useful sources of evidence about how effectively the Chapter 1 program meets its goals in performance and the duration of performance improvements. Major characteristics of portfolios include the following: (1) portfolios are a form of test and as such, theories and technologies of testing for reliability and validity can apply; (2) they can be useful and legitimate means of gathering data for the CHEPIS as they meet evaluation requirements, help participants draw situation-specific conclusions, and assist the participants in understanding students' learning better; (3) they have particular advantages for use in pre-kindergarten through first grade as they are specific and flexible for evaluating program progress toward desired outcomes; and (4) they have disadvantages as evaluation tools that must be recognized and controlled to the greatest extent possible. Statistical data are provided in two tables and one figure. Thirteen references are provided. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340806

PORTFOLIOS IN CHAPTER 1 EVALUATION AND PROGRAM IMPROVEMENT

ART BURKE

Paper presented at the Annual Meeting of the
Oregon Educational Research Association.
October 10-11, 1991
Portland, Oregon

Art Burke
Portland Public Schools
Evaluation Dept
PO Box 3107
Portland, OR 97208-3107

503 249-2000 x205

U0 028 478

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A Burke
Portland Pub Schools

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

This paper addresses methodological and substantive issues regarding the use of portfolios of students' work in Chapter 1 evaluation and program improvement. This paper is not a comprehensive methodological critique of the role of portfolios in evaluation. The purpose of the paper is to demonstrate that, with proper controls, portfolios can serve as an appropriate and important tool in Chapter 1 evaluation and program improvement, particularly in prekindergarten, kindergarten, and first grade programs. These programs are required to evaluate their services and implement program improvement when needed. However, they are not required to base evaluations on nationally-normed instruments. This paper presents practical suggestions for making portfolios serve the needs of Chapter 1 evaluation and program improvement within the specific requirements of Chapter 1 regulations.

This work is important for several reasons. First, portfolios are increasingly popular tools for assessment, especially in the early grades. Stiggins (1991) points out that traditional assessment methodology is poorly understood by school staffs. This emphasizes the importance of critical examination of alternative means of assessment such as portfolios. This paper discusses the methodological pitfalls which can trap uncritical users of portfolios and presents suggestions for making portfolios more valid assessment tools. Second, the specific use of portfolios in Chapter 1 evaluation and program improvement has not been addressed in the published literature. Third, portfolios can be important tools in integrating instruction and assessment. They can, for example, meet both the formal requirements of Chapter 1 evaluation and program improvement and the needs of classroom teachers for information to assess and guide students' learning.

This paper is organized into two parts. The first part discusses portfolios and methodological issues surrounding their use as evaluation tools. The second part presents the Chapter 1 Evaluation/Program Improvement System (CHEPIS) and approaches for making portfolios serve as technically sound instruments for use in CHEPIS. The Chapter 1 Policy Manual (U.S. Department of Education, 1990) is quoted and paraphrased throughout this paper without a citation for each reference.

PART ONE
PORTFOLIOS AND THE METHODOLOGY OF THEIR USE IN EVALUATION

Questions of what a portfolio is and how it should be constituted have been addressed by a number of authors (Wolf, 1989; Paulson, Paulson, and Meyer, 1991). Portfolios are finding use not only in schools, but in state-wide assessments in writing and math (Vermont Department of Education, 1991a, 1991b) and system-wide assessment in early childhood

programs (W. Spicer, personal communication, September 17, 1991). In this paper, a portfolio is defined, not in reference to a physical collection of materials, but as the collection of student work related to a single instructional theme. In other words, whatever the physical embodiment of a portfolio, what constitutes a portfolio for CHEPIS purposes is only the materials that will be subject to Chapter 1 evaluation.

DESIRABLE CHARACTERISTICS OF EVALUATION PROCEDURES

This section deals with the methodology of portfolio use in evaluation. We begin by briefly discussing characteristics desirable in all evaluation procedures. The discussion here is non-technical, more formal elaboration is given in Linn (1989).

Objectivity

An evaluation procedure is objective to the degree that different scorers of the same material obtain the same results. An example of a highly objective evaluation procedure is a multiple-choice standardized test with an answer sheet scored by computer. An example of a highly subjective evaluation procedure is the essay, where different raters may use different scoring rules.

Reliability

Reliability refers to the consistency of results produced by an evaluation method. All evaluation results are affected by measurement error. Measurement error is due to the presence of factors which affect test performance but are extraneous to the trait being measured. For example, a student's score on a mathematics test could be affected because the test contains material studied the night before, because the student was ill or fatigued at the time of testing, because the student hit a jackpot of lucky guesses, or because of a host of other circumstances not related to his or her real level of math achievement. Reliable evaluation results are those in which the influence of measurement error is minimized.

Validity

Validity refers to the adequacy and appropriateness of interpretations of the results of an evaluation method. For example, for a test to be a valid measure of achievement in history, scores have to depend on knowledge of history, not on reading ability or test-taking

skill. It is important to understand that validity is situational: An evaluation procedure may be valid for some purposes and invalid for others. For example, a test valid as a measure of achievement in third-grade history might not be valid as a predictor of achievement in fifth-grade history. Lastly, strictly speaking, it is interpretations or uses of evaluation results which are validated, not the evaluation procedure itself.

When portfolios of students' work are used as evaluation tools, ratings assigned to those portfolios are subject to examination of their objectivity, reliability, and validity in the same way as ratings from tests, inventories, checklists, questionnaires, observations, and other sources of information about students. As with other forms of assessment where students construct responses, portfolios may suffer from lower reliability and validity, require complicated scoring procedures, and involve difficulties with interrater agreement (Snow and Mandinach (1991). We will discuss specific threats to the reliability and validity of ratings of portfolios and specific suggestions for improving the reliability and validity of portfolio ratings later in this paper.

PART TWO
THE CHAPTER 1 EVALUATION/PROGRAM IMPROVEMENT SYSTEM

THE EVALUATION SYSTEM

Chapter 1 services are designed to improve the educational opportunities of educationally deprived[1] children by helping these children

1. Succeed in the regular school program;
2. Attain grade-level proficiency; and
3. Improve achievement in basic and more advanced skills.

The Chapter 1 Evaluation/Program Improvement System (CHEPIS) serves two purposes: to gather evidence on how successfully programs are meeting Chapter 1 goals, and, if goals are not met, to use evaluation results to improve programs. In order to appreciate how many uses portfolios can find in CHEPIS, we must examine CHEPIS in detail. Figure 1 presents the Chapter 1 Evaluation/Program Improvement System.

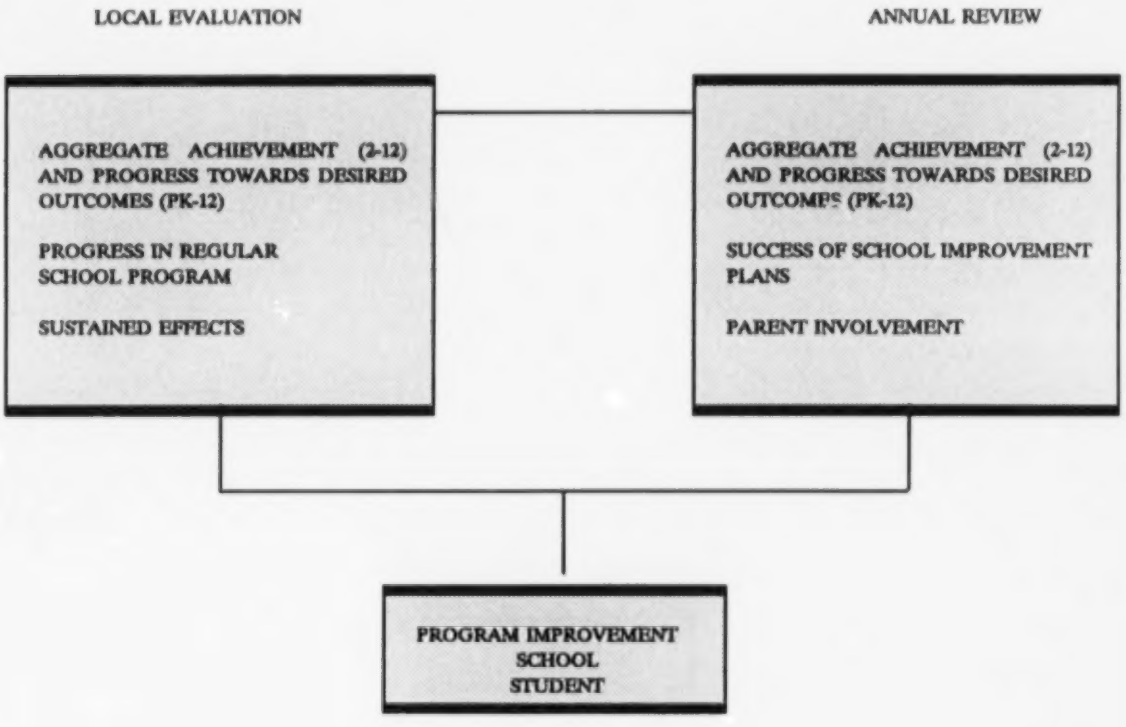
[INSERT FIGURE 1 ABOUT HERE]

---

[1] "Educationally deprived" refers to children whose educational attainment is below the level that is appropriate for their age.

FIGURE 1

THE CHAPTER 1 EVALUATION/PROGRAM IMPROVEMENT SYSTEM



CHEPIS is comprised of two subsystems: the Evaluation System and the Program Improvement System. The Evaluation System in turn has two components: the Local Evaluation, which is a three-year cycle of activities, and the Annual Review. The fundamental purpose of the Evaluation System is to provide evidence on how well the programs are meeting two standards: achievement gains on nationally-normed tests (Grades 2 - 12) and satisfactory progress towards the desired outcomes (PreK - 12). A major purpose of this paper is to specify ways in which progress towards desired outcomes can be effectively and validly evaluated by means of portfolios of students' work.

Desired Outcomes

A desired outcome is a goal statement or objective related to the purposes of Chapter 1. It focuses attention on what children will learn and accomplish as a result of participation in the Chapter 1 program. The statement of desired outcomes has a particular form in Chapter 1. It should specify the following:

> **Goal** -- What the children are to learn or accomplish.
> **Outcome Indicator** -- What will be used to measure achievement.
> **Standard or Performance Level** -- Specification of substantial progress
> **Time Frame** -- The period of time over which measurement will occur.

An example of a statement of a desired outcome in which portfolios are the outcome indicator is the following:

> Students in the first grade program will attain the writing and math skills necessary to advance to the second grade. Skills will be assessed by portfolios of students' work in writing and math. Satisfactory progress will occur if 75% of participating children receive a rating of "Satisfactory" or better for the work in their portfolio.

Chapter 1 projects are accountable for progress towards their desired outcomes -- when schools or even individual students fail to achieve substantial progress towards desired outcomes, the district may have to take steps to improve its Chapter 1 programs.

We will next briefly describe the essential features of the two components of the Chapter 1 Evaluation System -- the Local Evaluation and the Annual Review.

The Local Evaluation

At least once every three years, local educational agencies must evaluate the effectiveness of their Chapter 1 programs in two areas: the performance of Chapter 1 students in the regular school program and whether the beneficial effects of Chapter 1 are sustained over more than one program year.

Performance in the Regular School Program

One of the purposes of Chapter 1 is to help participants attain success in the regular school program. However, achievement gains produced in Chapter 1 programs may not translate into improved achievement in the regular school program. Conversely, significant progress by Chapter 1 participants in the regular program may not be captured by testing in the Chapter 1 program. As a result, local educational agencies must carefully assess the performance of Chapter 1 participants in the regular school program.

The basic requirements for assessing progress in the regular school program are as follows:

- Review Chapter 1 students' progress in the regular school program at least once every three years.
- Determine through the review whether the ultimate purpose of Chapter 1 -- success in the regular program -is being achieved.
- Include that information in the evaluation results reported to the State at least once every three years.

The areas covered by the assessment should reflect the shared goals and objectives between the regular program and Chapter 1. There should be a similarity of content taught, skills developed, and level of difficulty. The review may be based on teacher judgements, grade, retention rates, criterion-referenced tests, end-of-chapter or textbook tests, district or state administered tests, and other appropriate indicators of success, including portfolios of students' work. State and local officials determine the information to be collected and the method of collection.

Sustained Effects

The basic requirements of a sustained effects study include:

- ** Determining if the effect of the Chapter 1 program has been sustained over more than one program year;
- ** Assessment of the performance of the same students over two consecutive 12-month periods;
- ** Use of either a spring-spring-spring or a fall-fall-fall testing interval; and
- ** Use of results in planning program improvements.

Portfolios of students' work can be an important supplement to standardized, nationally normed tests in evaluating the sustained effects of Chapter 1 programs.

In conclusion, portfolios of students' work are useful sources of evidence about how effectively the Chapter 1 program meets its goals in two important areas: the performance of Chapter 1 students in the regular school program and whether the beneficial effects of Chapter 1 participation are sustained over more than one year.

The Local Review

The law requires that a LEA annually review whether the Chapter 1 project in each school improves student performance as measured by aggregate performance and the desired outcomes described in the LEA application. The purpose of the Local Review is to provide information to meet accountability requirements and for use in improving programs. Portfolios can be an effective means of providing evidence about how well the program is accomplishing its desired outcomes.

Pre-kindergarten, kindergarten, and first-grade projects are evaluated solely on the basis of their effectiveness in achieving their desired outcomes. For these project progress towards achieving desired outcomes may be measured

by improved student performance on CRTs, improved attendance, developmental checklists, observational scales, teacher ratings, skills mastery checklists, portfolios, and other means.

They are <u>not</u> required to use a fall-fall or spring-spring evaluation cycle.

They are <u>not</u> required to conduct studies of sustained effects.

Technical Standards For Chapter 1 Evaluation

Chapter 1 evaluations must meet prescribed technical standards. These standards are as follows:

(1) Evaluation findings must be representative of all students served.
(2) Evaluation instruments and procedures must be reliable and valid.
(3) Achievement gains must be addressed in a valid manner.
(4) Errors in evaluation procedures must be minimized through quality control mechanisms.

In a formal sense, these standards apply only to the use of nationally-normed tests in evaluating achievement in reading or mathematics in grades 2-12. However, they are technically relevant to the use of portfolios used with any grades as part of a Chapter 1 evaluation. The third standard relates entirely to the use of nationally-normed tests; it is not relevant to portfolio use and will not be discussed here. We will discuss the implications of portfolio use in relation to standards 1, 2, and 4.

Representativeness

Methods used in Chapter 1 evaluations must produce findings representative of all students served. Sampling procedures may be implemented for purposes of evaluating the success of Chapter 1 participants in the regular school program and of determining whether the beneficial effects of Chapter 1 participation are sustained over more than one year. However, progress towards the desired outcomes must be assessed for each Chapter 1 participant each year. If use of portfolios causes results for many students to be omitted from the annual review, the representativeness requirement would be violated. An example of the kinds of difficulties which may develop around using portfolios comes from Vermont's Math Portfolio Assessment Program, where only 83% of the 4th grade portfolios, and 58% of the 176 eighth-grade portfolios were scorable (Vermont Department

of Education, 1991b). The conclusion was that the state needs to be more definitive and more prescriptive in terms of portfolio content before the next statewide assessment.

Quality Control

Chapter 1 regulations require that care be taken with the evaluation procedures in order to produce the greatest number of accurate and valid scores. In short, procedures have to be implemented to reduce errors computing, converting, and recording scores. Mechanisms to minimize such errors include repeating procedures on a random sample of scores or calculations, establishing checks for consistency and accuracy in evaluation results, and providing training and/or tutorial materials to staff.

Reliable and Valid Measurement

Reliability and validity have been discussed in general terms earlier. For purposes of Chapter 1 evaluation, "curricular" validity is essential; that is, the degree of correspondence between the curriculum or instructional content of the Chapter 1 program and the content of the test. Table 1 presents questions to orient users towards the curricular validity of portfolios used in Chapter 1 evaluation. Ways of improving the reliability and validity of ratings of portfolios used in Chapter 1 evaluation will be discussed in the section headed "Technical Standards for Portfolio Use in CHEPIS."

TABLE 1
THE CURRICULAR VALIDITY OF A PORTFOLIO

DOES THE CONTENT AND DIFFICULTY OF THE PORTFOLIO MATCH WHAT WAS ACTUALLY TAUGHT?
DOES CONTENT AND DIFFICULTY OF THE PORTFOLIO MATCH THE CURRICULAR GOALS EXPRESSED IN THE STATEMENT OF DESIRED OUTCOME?

THE PROGRAM IMPROVEMENT SYSTEM

School Program Improvement

The law requires an LEA to implement a formal process of program improvement for each school not showing substantial progress[2] toward meeting the desired outcomes described in the application or showing no improvement or a decline[3] in aggregate performance for a 12-month period.

A LEA must assess whether Chapter 1 students in each school as a whole are showing improvement. If not, the LEA is required to take action until improvement is shown. The results of the local evaluation and the local review must be considered in planning school improvement programs.

Student Program Improvement

With regard to individual students, the local educational agency must use the results of the local evaluation and the local review to do the following:

1. Identify all children who have been served for a school year and have not shown substantial progress toward meeting the desired outcomes or whose performance shows no improvement or a decline[4]. Modifications in the chapter 1 project must be considered to serve those students better

2. Conduct a thorough assessment of the educational needs of children who have not shown substantial progress toward meeting the desired outcomes or whose performance shows no

---

[2] "Substantial progress" is defined by the LEA.

[3] "No improvement or a decline" occurs if Chapter 1 participants fail to make gains beyond that which they would be expected to make in the absence of the additional help the program provided. The measure of this is the NCE criterion set (in Oregon) by the State Department of Education.

[4]The definition of "no improvement or a decline" is the same for individual students as it is for schools. See footnote 3.

> improvement or a decline after two consecutive years in the project. If appropriate, use the results of the assessment to modify the Chapter 1 project to meet the children's needs. The LEA is not required to develop an individual educational plan for each student.

When portfolios are used as a tool to evaluate progress towards desired outcomes, the results have to be used in two ways: to identify students who may require program improvement and to provide evidence suggestive of appropriate improvements. A caution is necessary. Portfolios useful for evaluation may not be useful for suggesting the kinds of improvements to programs required under these regulations. If portfolios are to be useful in diagnosis, objective scoring or classification systems should be built into the tasks and explicit decision points specified for instructional steps (Snow and Mandinach, 1991). Burke (1991) has proposed an Assessment/Action model for Chapter 1 Student Program Improvement which includes specific suggestions for using evaluation results to improve Chapter 1 instructional programs for individual students.

TECHNICAL STANDARDS FOR PORTFOLIO USE IN CHEPIS

Table 2 lists some of the sources of error which may threaten the reliability and validity of ratings applied to the contents of student portfolios. Information in Table 2 was compiled from methodological analyses of rating methods (Gronlund and Linn, 1990; Ebel and Frisbie, 1991; and Evertson and Green, 1986).

[INSERT TABLE 2 ABOUT HERE}

Table 2 makes clear that in order for portfolios to be useful assessment tools, care must be taken to avoid the many threats to their reliability and validity. The following suggestions, based on discussion in Gronlund and Linn (1990) about improving the quality of ratings of students' work, presents ways of improving the objectivity, reliability, and validity of ratings applied to portfolios to make those ratings conform as far as possible to the technical standards for Chapter 1 evaluation.

1. **Limit portfolio content to material directly related to the desired outcomes.** The contents and difficulty of material in the portfolios must reflect what was actually taught. Many teachers allow students to select work for inclusion in their portfolio. This practice is a threat to the curricular validity of using portfolios in Chapter 1 evaluations if the selected material is not representative

TABLE 2
SOURCES OF ERROR IN RATING STUDENT PORTFOLIOS

Type of Error	Definition
1. Central tendency	Tendency to avoid extreme ratings -- rate everyone towards the scale midpoint.
2. Leniency or generosity	Tendency to assign favorable ratings.
3. Severity	Tendency to assign unfavorable ratings.
4. Halo effect	General impression of a person influences the rating of individual characteristics.
5. Observer interference	The observer becomes an actor in the setting or the performance.
6. Logical error	Two characteristics are rated as similar because of the rater's beliefs concerning their relationship (i.e. gifted children are poorly adjusted socially).
7. Subjectivity	"Looking with blinders" and "watching without seeing -- selective looking, (i.e. ignoring the mistakes of a "good" student.
8. Inference	The reason for a particular behavior is not always apparent or sometimes apparent reasons are incorrect.
9. Trait ambiguity	If the characteristic to be observed is described in general, implicit terms, considerable inference will be required to detect instances of it. e.g. "off task"
10. Subject consistency	Judgments may be based on evidence from an unrepresentative sample.
11. Classification	Macro categories hide fine distinctions.
12. Bias	Distortions due to unchecked personal bias of observer.

13. Sampling	Samples do not represent the general group of behaviors, that do not occur frequently enough, or are inconsistent with the theory guiding the observations.
14. Reactivity	Reactions of participants being observed can distort the process or phenomenon being observed.
15. Situation or context	False assumptions of functional equivalence (e.g., testing time 1 = testing time 2).
16. Inadequate rating system	Compromises reliability and validity.
17. Simultaneity	Failure to account for more than one activity occurring at a time; more than one message sent at a time; and more than one function of a message.
18. Drift	Changes in the way ratings are assigned over time.
19. Primacy or recency	First impressions distort later judgments.

of what was actually taught. Furthermore, the content and difficulty of the material in the portfolios should fit closely the statements of desired outcomes.

2. **Rate characteristics of portfolios directly related to the desired outcomes.** These have to include, by definition, the basic and more advanced skills that all children of a particular age or grade level are expected to master. If portfolios are used as part of the Chapter 1 evaluation for particular grades at different schools, the same characteristics should be rated at all schools using portfolios.

3. **Characteristics and points on the rating scale should be clearly defined.** This will be illustrated below for a desired outcome expressed in terms of mathematical problem solving.

    1. To what extent does the material in the student's portfolio demonstrate sound mathematical problem solving?

        1. Completely inappropriate or unworkable strategies
        2. Mostly inappropriate or unworkable strategies
        3. Occasionally appropriate or workable strategies
        4. Consistently workable approaches
        5. Efficient or sophisticated approaches in evidence

4. **Portfolios should be rated according to uniform procedures.** Uniform procedures should be developed for scoring the portfolios and recording and reporting the results. Staff may require training in rating the portfolios. Ideally, the format for reporting results should be standardized across children, classrooms, and schools.

5. **Whenever possible, ratings of several observers should be combined.** Pooled ratings of several teachers will generally be more reliable than ratings obtained from any one teacher. When ratings are combined by a process such as averaging, personal biases and other types of errors listed in Table 1 will tend to cancel one another out.

SUMMARY REMARKS ABOUT PORTFOLIOS IN CHAPTER 1 EVALUATION AND PROGRAM IMPROVEMENT

1. **A test is a test is a test ..... A portfolio is in many ways a test.**

   --- The technology of testing is well-established and has a basis in highly evolved theories of reliability and validity and procedures for estimating the reliability and validity of measuring instruments in specific situations. These ideas are useful in analyzing the appropriateness of using portfolios as evaluation tools.

   --- The validity of judgements or scores based on portfolio methods should be constantly evaluated (Messick, 1990).

2. **Portfolios can be a useful and legitimate means of gathering data for purposes of Chapter 1 Evaluation and Program Improvement.**

   --- Portfolios can meet both Chapter 1 evaluation requirements and the needs of classroom teachers for samples of students' work. As a result, they may be very useful in improving coordination between the Chapter 1 and regular programs.

   --- An advantage of portfolios over other evaluation tools is that portfolios can help students, teachers, and parents draw conclusions about their own specific situations.

   --- The real test of the usefulness of portfolios is whether they help students better understand their learning, parents better understand their children's learning in relation to the curriculum, teachers better rate their pupils' learning and plan for next steps.

   --- Portfolios may be more useful for assessing progress towards curricular goals and less useful for reliable

comparisons among students, particularly when students control the content of the portfolio.

--- The portfolio is a tool of evaluation, not an instructional end in itself -- it should not become the curriculum.

3. **Portfolios have particular advantages for use in CHEPIS from pre-kindergarten through first grade.**

--- These programs are evaluated solely on their progress towards their desired outcomes. Portfolios are both specific and flexible means of evaluating progress towards desired outcomes.

4. **The disadvantages of portfolios as evaluation tools must be recognized and controlled as far as possible.**

--- Disadvantages include problems with objectivity, scoring, reliability, validity, and so on. With proper controls, threats to the reliability and validity of evaluation of portfolios can be lessened.

REFERENCES

Burke, A. (1991, October). Implementing Chapter 1 student program improvement in state and local educational agencies. Paper presented at the annual meeting of the Oregon Educational Research Association, Portland, Oregon.

Ebel, R. L., & Frisbie, D. A. (1991). Essentials of educational measurement (5th ed.). Engelwood Cliffs, New Jersey: Prentice Hall.

Evertson, C. M., & Green, J. L. (1986). Observation as inquiry and method. In M. C. Wittrock (Ed.). Handbook of research on teaching (3rd ed.). pp 162-213. New York: MacMillan Publishing Company.

Gronlund, N. E., & Linn, R. L. (1990). Measurement and evaluation in teaching (6th ed.). New York: MacMillan Publishing Company.

Linn, R. L. (Ed.) (1989). Educational measurement (3rd ed.). New York: MacMillan Publishing Company.

Messick, S. (1990). Validity of test interpretation and use. Princeton: Educational Testing Service.

Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? Educational Leadership, 48, 5, 60-63.

Snow, R.E., & Mandinach, E.B. (1991). Integrating assessment and instruction: A research and development agenda, Princeton, NJ: Educational Testing Service.

Stiggins, R. J. (1991). Assessment literacy. Phi Delta Kappan, 72, 7, 534-539.

U.S. Department of Education. (1990). Chapter 1 policy manual. Washington, DC: Author.

Vermont Department of Education. (1991a). This is my best: The report of Vermont's writing assessment program. Montpelier, VT: Author.

Vermont Department of Education. (1991b). Looking beyond "The Answer: The report of Vermont's mathematics portfolio assessment program. Montpelier, VT: Author.

Wolf, D. P. (1989). Portfolio assessment: Sampling student work. Educational Leadership, 46, 7, 35-39.